What is claimed is:

1. A method for factoring an input finite-state transducer (FST) including an unknown symbol, comprising the steps of:

replacing each occurrence of the unknown symbol in the input FST with the unknown symbol and a diacritic to define a left-sequential finite-state transducer (FST); and

replacing each occurrence of the diacritic with a symbol representative of an empty string and an output symbol to define a right-sequential finite-state transducer (FST);

wherein said replacing steps avoid direct factorization of the unknown symbol.

2. The method of claim 1, further comprising the step of factoring the unknown symbol in the input FST into arc label sequences $\lceil ?,\ \delta{:}\lambda_i \rceil_{LR}$ and $\lceil \lambda_i{:}\varepsilon,\ ?{:}\sigma^{out} \rceil_{RL}$, where:

λi is a diacritic,

σ^{out} is an output symbol, and

δ is a deterministic empty string.

3. The method of claim 2, further comprising the step of copying the arc label sequence $\lceil ?,\ \delta{:}\lambda_i \rceil_{LR}$ to the left-sequential FST.

4. The method of claim 2, further comprising the step of copying the arc label sequence $\lceil \lambda_i{:}\varepsilon,\ ?{:}\sigma^{out} \rceil_{RL}$ to the right-sequential FST.

5. The method of claim 1, wherein the left-sequential FST and the right-sequential FST are adapted for performing language processing.

6. The method of claim 5, wherein the language processing comprises one of tokenization, phonological analysis, morphological analysis, disambiguation, spelling correction, and shallow parsing.

7. The method of claim 1, wherein the left-sequential FST and the right-sequential FST are lexical transducers.

8. An apparatus for factoring an input finite-state transducer (FST) including an unknown symbol, comprising:

means for replacing each occurrence of the unknown symbol in the input FST with the unknown symbol and a diacritic to define a left-sequential finite-state transducer (FST); and

means for replacing each occurrence of the diacritic with a symbol representative of an empty string and an output symbol to define a right-sequential finite-state transducer (FST);

wherein said replacing means avoid direct factorization of the unknown symbol.

9. The apparatus of claim 8, further comprising means for factoring the unknown symbol in the input FST into arc label sequences $\lceil ?, \delta{:}\lambda_i \rceil_{LR}$ and $\lceil \lambda_i{:}\varepsilon, ?{:}\sigma^{out} \rceil_{RL}$, where:

λi is a diacritic,

σ^{out} is an output symbol, and

δ is a deterministic empty string.

10. The apparatus of claim 9, further comprising means for copying the arc label sequence $\lceil ?, \delta{:}\lambda_i \rceil_{LR}$ to the left-sequential FST.

11. The apparatus of claim 9, further comprising means for copying the arc label sequence $\lceil \lambda_i{:}\varepsilon, ?{:}\sigma^{out} \rceil_{RL}$ to the right-sequential FST.

12. The apparatus of claim 8, wherein the left-sequential FST and the right-sequential FST are adapted for performing language processing.

13. The apparatus of claim 12, wherein the language processing comprises one of tokenization, phonological analysis, morphological analysis, disambiguation, spelling correction, and shallow parsing.

14. The apparatus of claim 8, wherein the left-sequential FST and the right-sequential FST are lexical transducers.